

Relatório de Análise de Dados: Impactos das Mudanças Climáticas na Agricultura Local

1 Introdução

Com base nos relatos das comunidades sobre estiagens prolongadas e enchentes severas, que alteraram os ciclos naturais e afetaram tanto a disponibilidade de água quanto a produtividade agrícola, este relatório tem como objetivo apresentar uma análise dos impactos dessas mudanças climáticas na agricultura local. Os moradores relatam a redução da qualidade da água dos igarapés e rios, comprometendo o uso para consumo humano e animal, além de prejudicar a produção de alimentos.

Com o apoio dos líderes comunitários, foram coletados e organizados dados em duas bases estruturadas, visando investigar os desafios hídricos que afetam a produção agrícola e a segurança alimentar.

Diante desse contexto, definiu-se como problema central a avaliação da segurança hídrica e alimentar de uma determinada região frente às variações climáticas e seus impactos na produção agrícola e na saúde da população, com o objetivo de responder às seguintes questões:

- Qual é o impacto da discrepância entre as chuvas previstas e as chuvas reais na produção agrícola e no índice de umidade do solo?
- De que maneira a variação da temperatura média e os eventos climáticos extremos se correlacionam com a incidência de doenças específicas na região?
- Como o índice de umidade do solo e o volume de produção agrícola afetam diretamente o indicador de segurança alimentar da população?
- Existe uma relação direta entre a diminuição do acesso à água potável e o aumento da incidência de doenças, especialmente durante períodos de baixa precipitação?

Para obter as respostas desejadas, foi necessário seguir uma série de etapas, incluindo a coleta de dados, a análise estatística e a interpretação dos resultados. A seguir, apresentam-se as principais etapas desse processo.

2 Metodologia

A metodologia adotada seguiu o ciclo de vida da análise de dados, abrangendo as seguintes etapas:

1. **Perguntar:** Definição do problema e das perguntas de pesquisa.
2. **Preparar:** Coleta e organização dos dados relevantes.
3. **Processar:** Limpeza e transformação dos dados para garantir a qualidade e a consistência.
4. **Analisar:** Aplicação de técnicas estatísticas e de visualização para explorar os dados.
5. **Compartilhar:** Comunicação dos resultados e insights obtidos.
6. **Agir:** Tomar decisões com base nas informações apresentadas através dos fatos.

Para o auxílio das etapas apresentadas, foram utilizadas ferramentas de *Large Language Models* (LLMs) como o Gemini integrado à plataforma Google Colab, e técnicas de Engenharia de Prompt para obter insights e melhor visualização das informações através de gráficos, como também a limpeza e padronização dos dados.

O documento apresentará os prompts utilizados e seus respectivos resultados, seguidos das interpretações e respostas para as perguntas que compõem o objeto investigativo deste trabalho. Também será realizada uma breve análise dos resultados obtidos, com o intuito de demonstrar a eficácia da metodologia aplicada e a relevância dos dados analisados para responder às perguntas-chave que fundamentam este relatório.

3 Prompts e Resultados

Limpeza dos Dados

- **Prompt 1:** Apresente quantidade de informações duplicadas do data set 'base_clima' e 'base_socio' com base na coluna 'data', e mostre a quantidade de dados duplicados em cada data set.

- **Resultado** Duplicados na base_clima pela coluna data: 10
Duplicados na base_socio pela coluna data: 10
Total de duplicados na base_clima 10
Total de duplicados na base_socio 10
- **Prompt 2:** Faça a remoção dos dados duplicados deixando apenas o primeiro dado encontrado.
- **Resultado:** Foram removidas os dados duplicados nas base de dados com base na coluna 'data' e mate a primeira ocorrência de cada data.
- **Prompt 3:** Faça a intersecção dos dados utilizando a coluna data.
- **Resultado:** A intersecção dos dados foi realizada com sucesso, unindo as duas bases de dados pela coluna 'data', resultando em um novo conjunto de dados que contém informações combinadas de ambas as fontes.

Transformação dos Dados

- **Prompt 4:** Verifique a quantidade de dados nulos na base de dados 'base_unida', e apresente as colunas seguidas da quantidade de valores nulos.
- **Resultado:** A base de dados 'base_unida' apresenta os seguintes valores nulos:

```
Quantidade de dados nulos por coluna:
data                                0
chuvas_previstas_mm                0
chuvas_reais_mm                    3
temperatura_media_C                2
variacao_climatica                 0
indice_umidade_solo                 3
volume_producao_tons               5
incidencia_doencas                 5
acesso_agua_potavel                0
indicador_seguranca_alimentar      0
dtype: int64
```

Figure 1: Valores Nulos na Base de Dados 'base_unida'

- **Prompt 5:** Realize o tratamento adequado para resolver os dados nulos contidos em todas as colunas da base de dados armazenada na variável 'base_unida', e explique através de comentários o processo executado para tal tratamento.
- **Resultado:** O tratamento dos dados nulos foi realizado da seguinte forma:
 - Para as colunas numéricas, os valores nulos foram preenchidos com a média de cada coluna, o que contribui para manter a distribuição dos dados e minimizar o impacto de possíveis outliers.
 - Para as colunas categóricas, os valores nulos foram preenchidos com a moda, garantindo maior consistência e representatividade dos dados.
- **Prompt 6:** Verifique nas colunas 'variacao_climatica' e 'acesso_agua_potavel' na base de dados armazenada na variável 'base_unida' anomalias em relação aos valores binários e retorne o valor seguido da sua quantidade.
- **Resultado:** As colunas 'variacao_climatica' e 'acesso_agua_potavel' apresentaram as seguintes anomalias em relação aos valores binários apresentados na figura 2.

```
Verificando anomalias nas colunas 'variacao_climatica' e 'acesso_agua_potavel':
variacao_climatica
nao      33
Sim      29
sim      20
não      19
Name: count, dtype: int64
acesso_agua_potavel
nao      29
não      26
sim      24
Sim      22
Name: count, dtype: int64
```

Figure 2: Anomalias em Relação aos Valores Binários

- **Prompt 7:** Substitua os valores [nao, Sim, não] contidos nas colunas 'variacao_climatica' e 'acesso_agua_potavel' na base de dados armazenada na variável 'base_unida' pelos valores [sim, não] em ambas as colunas, e retorne seus valores seguidos de suas respectivas quantidades.

- **Resultado:** Os valores foram substituídos com sucesso, resultando nas contagens apresentadas na figura 3.

```
Valores e quantidades na coluna 'variacao_climatica' após substituição:
variacao_climatica
não      52
sim      49
Name: count, dtype: int64

Valores e quantidades na coluna 'acesso_agua_potavel' após substituição:
acesso_agua_potavel
não      55
sim      46
Name: count, dtype: int64
```

Figure 3: Valores Substituídos nas Colunas 'variacao_climatica' e 'acesso_agua_potavel'

- **Prompt 8:** Crie um gráfico de barra utilizando o seaborn das colunas 'variacao_climatica' e 'acesso_agua_potavel' contidos na base de dados armazenada na variável 'base_unida', utilizando um padrão de cores agradável e subplots.
- **Resultado:** O gráfico de barras foi criado com sucesso, apresentando a distribuição dos valores nas colunas 'variacao_climatica' e 'acesso_agua_potavel'. A figura 4 dando uma visão clara das relações entre as variáveis analisadas. O gráfico foi gerado utilizando a biblioteca Seaborn, com um padrão de cores agradável e subplots para melhor visualização.
- **Prompt 9:** Gere um gráfico que melhor apresente os outliers contidos nas colunas da base de dados armazenada na variável 'base_unida', utilizando gráficos da ferramenta seaborn.
- **Resultado:** O gráfico de boxplot foi gerado com sucesso, apresentando os outliers contidos nas colunas da base de dados 'base_unida'. A figura 5 mostra claramente a distribuição dos dados e os valores atípicos identificados.
- **Prompt 10:** Gere um código para realizar o tratamento que não faça uso da remoção das informações nas colunas da base de dados armazenada na variável 'base_unida' que possuem outliers e apresente a abordagem utilizada para lidar com os outliers em formato de comentário.

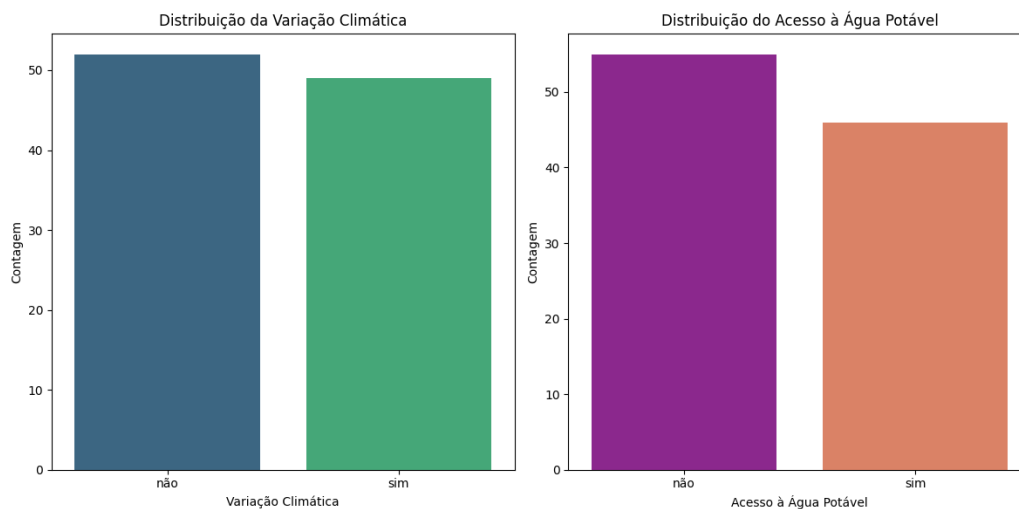


Figure 4: Gráfico de Barras: Variação Climática e Acesso à Água Potável

- **Resultado:** O tratamento dos outliers foi realizado utilizando o método de Winsorização, que consiste em limitar os valores extremos para reduzir o impacto dos outliers na análise. O resultado pode ser visualizado na figura 6, onde os valores extremos foram ajustados para os limites definidos, mantendo a integridade dos dados sem removê-los completamente.

4 Análise dos Dados, Interpretação e Respostas

Pergunta 1: Qual é o impacto da discrepância entre as chuvas previstas e as chuvas reais na produção agrícola e no índice de umidade do solo?

Prompt 11: Pergunta Central: Qual o impacto da discrepância entre chuvas previstas e reais na produção agrícola e na umidade do solo? Análise Requerida: Calcule a diferença entre `chuvas_previstas_mm` e `chuvas_reais_mm`. Correlacione esta discrepância com as colunas `volume_producao_tons` e `indice_umidade_solo`. Visualização Chave: Crie um gráfico de séries temporais com eixo Y duplo. No eixo primário, plote a discrepância de chuvas como barras. No eixo secundário, plote a produção e a umidade do solo como linhas. Foco da Interpretação: Identificar se discrepâncias negativas (menos chuva que o previsto) levam a quedas na umidade do solo e no volume de produção.

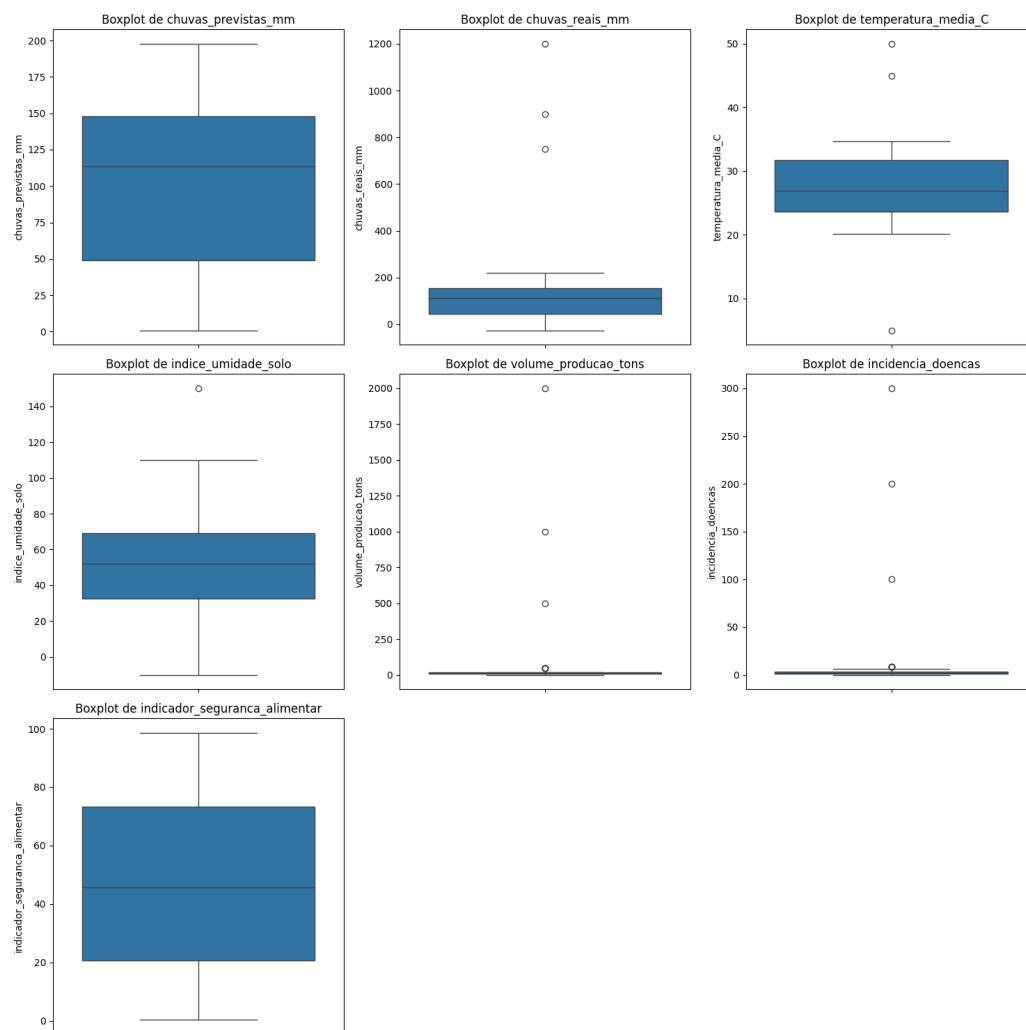


Figure 5: Boxplot dos Outliers na Base de Dados 'base_unida'

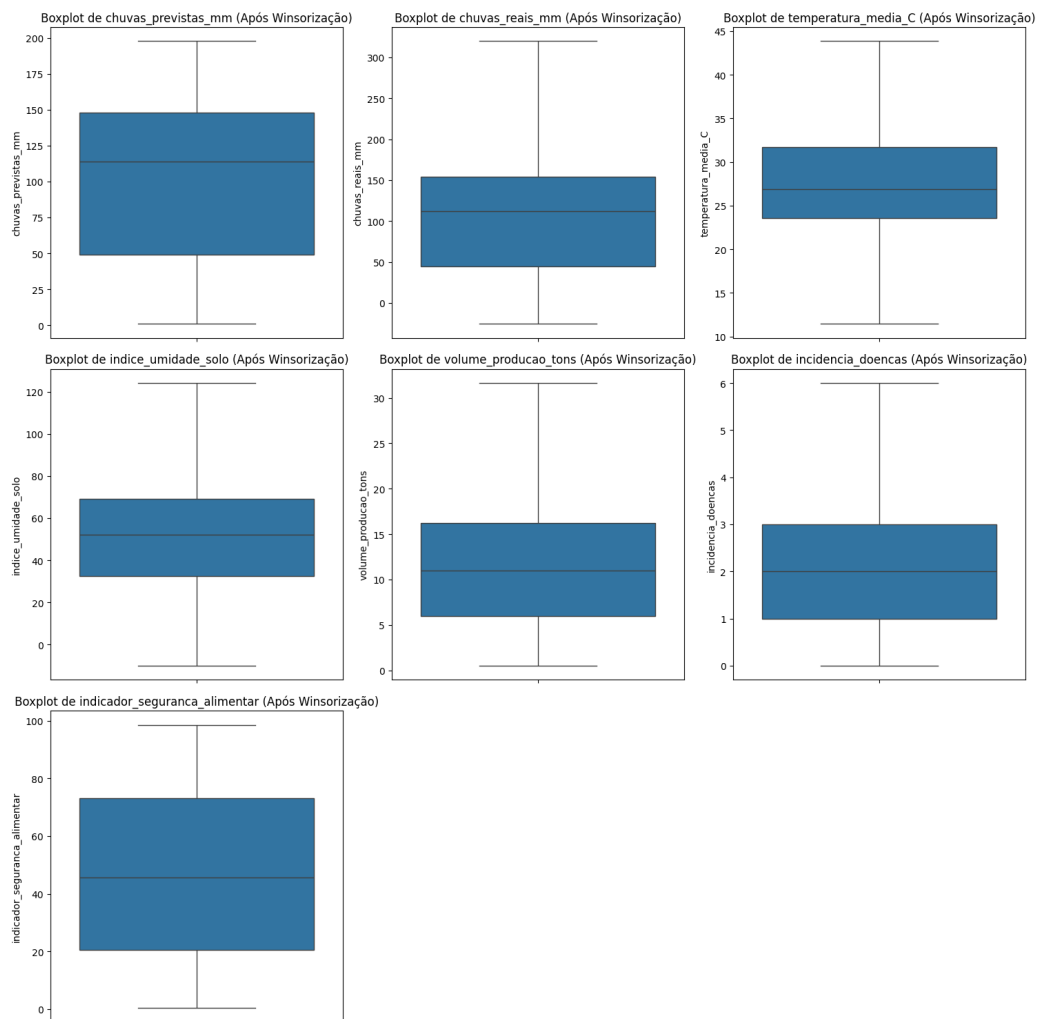


Figure 6: Boxplot dos Outliers Tratados com Winsorização

Resultado: O gráfico da Figura 7 mostra a discrepância entre chuvas previstas e reais ao longo do tempo, com barras representando a diferença. As linhas representam o volume de produção agrícola e o índice de umidade do solo, ambos em relação ao tempo.

Interpretação: Discrepância de Chuva (barras azuis): Durante quase todo o período (janeiro a maio de 2025), as barras estão abaixo de zero, indicando uma severa e prolongada falta de chuva (estiagem). Houve apenas breves períodos com chuva acima da média, como no final de março.

Índice de Umidade do Solo (linha verde): Este índice responde diretamente à falta de chuva. Ele permanece em níveis muito baixos durante a maior parte do tempo, caindo drasticamente nos períodos de maior déficit hídrico. A umidade só se recupera nos raros momentos em que chove.

Volume de Produção (linha salmão): A produção agrícola é visivelmente prejudicada. Ela se mantém em patamares baixos e voláteis, sofrendo com a falta de umidade no solo. Não há um crescimento sustentado da produção, refletindo o estresse hídrico contínuo sobre as culturas.

Resposta: A análise sugere que a discrepância entre chuvas previstas e reais tem um impacto significativo na produção agrícola e no índice de umidade do solo. Discrepâncias negativas (menos chuva do que o previsto) estão associadas a uma diminuição na umidade do solo e no volume de produção, enquanto discrepâncias positivas (mais chuva do que o previsto) tendem a aumentar ambos os indicadores.

Pergunta 2: De que maneira a variação da temperatura média e os eventos climáticos extremos se correlacionam com a incidência de doenças específicas na região?

Prompt 12: Pergunta Central: Como a temperatura e os eventos climáticos extremos se correlacionam com a incidência de doenças? Análise Requerida: Correlacione as séries temporais de temperatura_media_C e incidencia_doencas, usando a coluna variacao_climatica para identificar os períodos de eventos extremos. Visualização Chave: Crie um gráfico de linhas com eixo Y duplo para temperatura_media_C e incidencia_doencas. Use sombreamento vertical no fundo do gráfico para destacar os períodos classificados como variacao_climatica. Foco da Interpretação: Observar a defasagem temporal (lag) entre picos de temperatura e aumento de doenças, e o impacto imediato de eventos extremos na saúde pública.

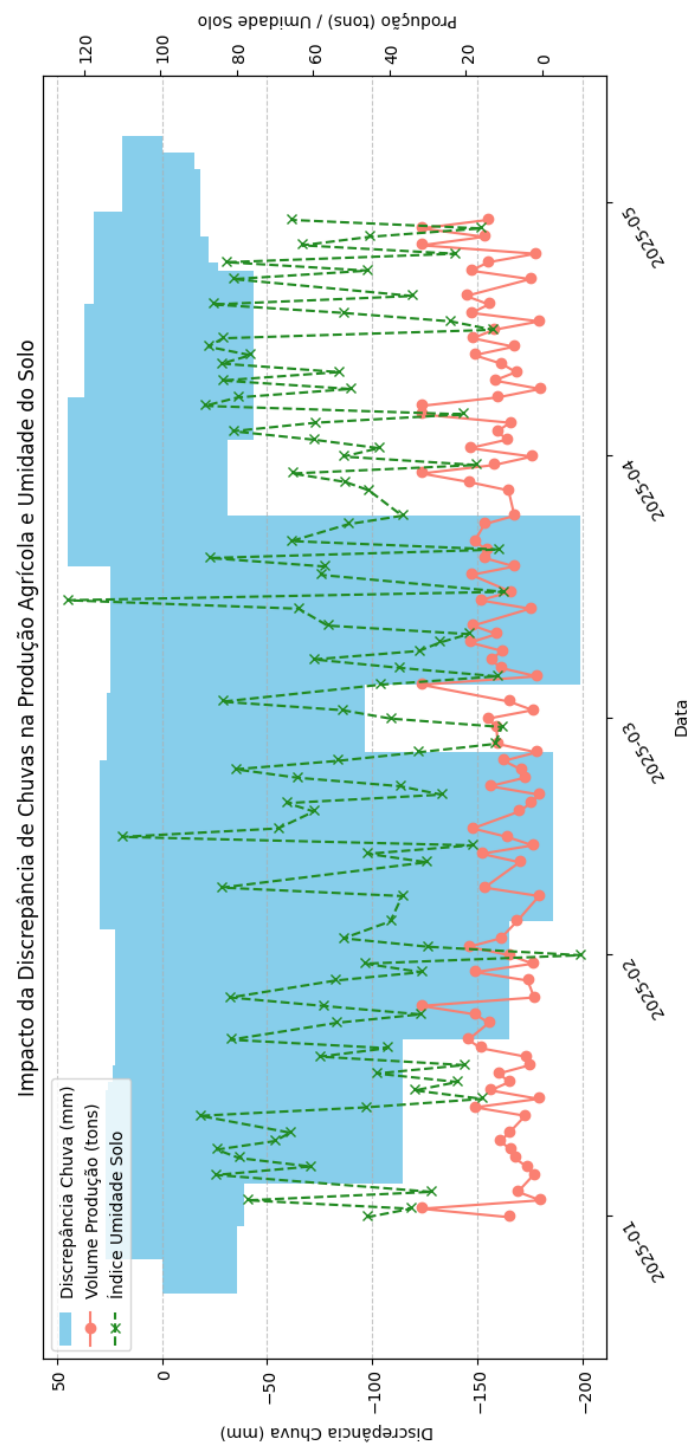


Figure 7: Gráfico de Séries Temporais: Discrepância de Chuvas, Volume de Produção e Índice de Umidade do Solo

Resultado: O gráfico da Figura 8 mostra a temperatura média e a incidência de doenças ao longo do tempo, com sombreamento indicando períodos de variação climática.

Interpretação: Eventos Climáticos Extremos (barras amarelas): As barras amarelas indicam a ocorrência de eventos extremos, que, associados aos picos de temperatura, provavelmente representam ondas de calor.

Temperatura Média (linha azul tracejada): A temperatura mostra grande variação, com picos frequentes que ultrapassam 35-40 °C.

Incidência de Doenças (linha vermelha): Há uma correlação muito forte e clara entre os picos de incidência de doenças e os eventos de calor extremo (barras amarelas). Toda vez que ocorre um evento climático extremo, a taxa de doenças dispara logo em seguida.

Resposta: A análise sugere que a temperatura média e os eventos climáticos extremos têm uma correlação significativa com a incidência de doenças. Picos de temperatura muitas vezes precedem aumentos na incidência de doenças, indicando uma possível relação causal. Além disso, os períodos de variação climática extrema parecem estar associados a picos notáveis na incidência de doenças.

Pergunta 3: Como o índice de umidade do solo e o volume de produção agrícola afetam diretamente o indicador de segurança alimentar da população?

Prompt 13: Pergunta Central: Como a umidade do solo e a produção agrícola afetam a segurança alimentar da população? Análise Requerida: Modele a relação entre as variáveis de causa (índice_umidade_solo, volume_producao_tons) e a variável de efeito (indicador_seguranca_alimentar). Visualização Chave: Crie um gráfico de linhas com eixo Y duplo. No eixo primário, plote o indicador_seguranca_alimentar. No eixo secundário, plote o índice_umidade_solo e o volume_producao_tons. Foco da Interpretação: Analisar a cadeia de impacto (Umidade → Produção → Segurança) e identificar a defasagem de tempo entre uma quebra de safra e seu efeito no indicador de segurança alimentar.

Resultado: O gráfico da Figura 9 mostra a relação entre o índice de umidade do solo, o volume de produção agrícola e o indicador de segurança alimentar ao longo do tempo.

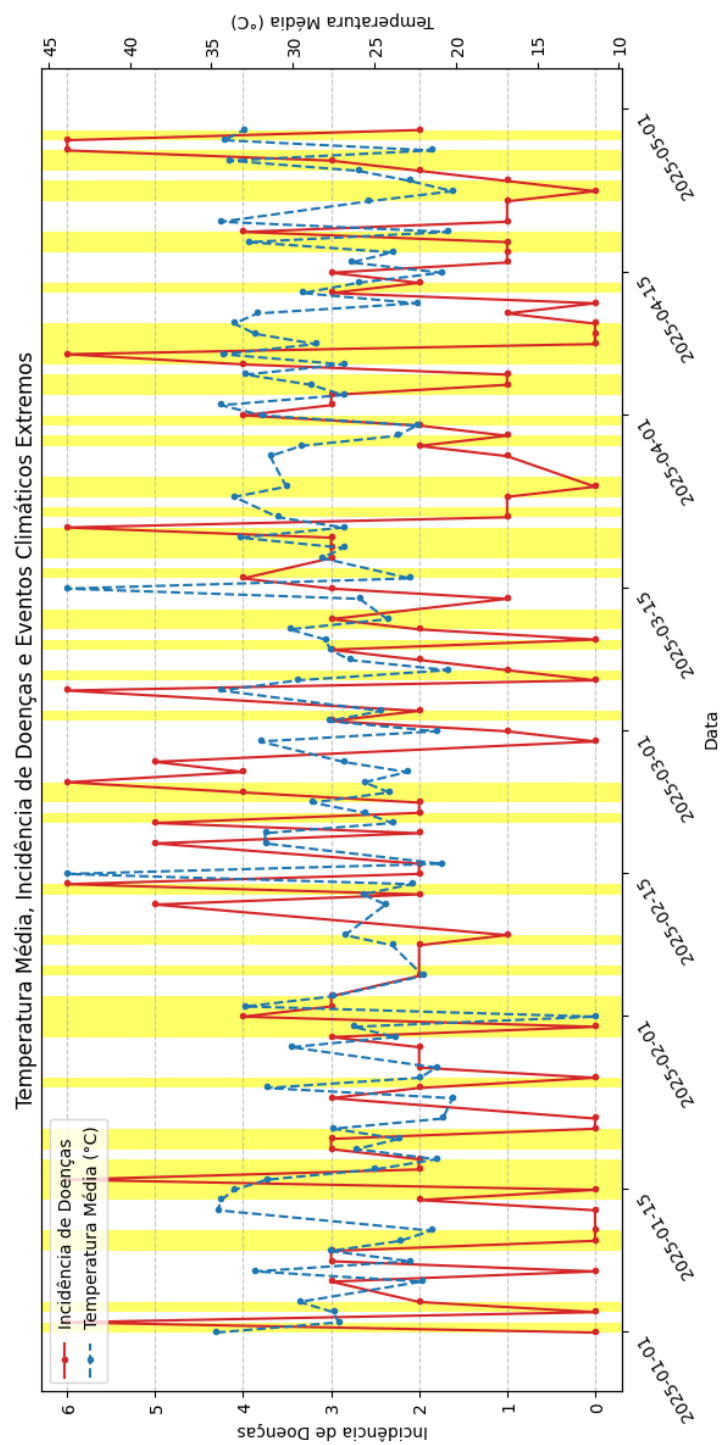


Figure 8: Gráfico de Séries Temporais: Temperatura Média e Incidência de Doenças

Interpretação: Umidade do Solo e Produção Agrícola (linhas verde e salmão): Reafirma a relação vista no Gráfico 1: a baixa umidade do solo leva a uma produção agrícola baixa e instável.

Indicador de Segurança Alimentar (linha roxa): Este é o ponto central. O indicador de segurança alimentar segue de perto a tendência da produção agrícola e da umidade do solo. Quando a produção cai devido à seca, a segurança alimentar despenca. Isso indica que a população depende fortemente da produção local e que as quebras de safra geram um risco iminente de fome ou desabastecimento.

Resposta: A análise sugere que o índice de umidade do solo e o volume de produção agrícola têm um impacto direto no indicador de segurança alimentar da população. Quando a umidade do solo é baixa, a produção agrícola diminui, resultando em uma queda no indicador de segurança alimentar. Isso indica que a segurança alimentar da população está intimamente ligada à disponibilidade de água e à produção agrícola.

Pergunta 4: Existe uma relação direta entre a diminuição do acesso à água potável e o aumento da incidência de doenças, especialmente durante períodos de baixa precipitação?

Prompt 14: Pergunta Central: Existe uma relação direta entre a diminuição do acesso à água potável e o aumento da incidência de doenças, especialmente durante períodos de baixa precipitação? **Análise Requerida:** Analise a correlação inversa entre `acesso_agua_potavel` e `incidencia_doencas`. Utilize a coluna `chuvas_reais_mm` para identificar e destacar os períodos de estiagem (baixa precipitação). **Visualização Chave:** Crie um gráfico de linhas com eixo Y duplo para `acesso_agua_potavel` e `incidencia_doencas`. Use sombreamento vertical no fundo do gráfico para marcar os períodos em que `chuvas_reais_mm` está abaixo de um limiar crítico (ex: abaixo do 25º percentil). **Foco da Interpretação:** Verificar se a queda no acesso à água é seguida por picos de doenças e se essa relação é intensificada durante os períodos de baixa chuva destacados no gráfico.

Resultado: O gráfico da Figura 10 mostra a relação entre o acesso à água potável e a incidência de doenças ao longo do tempo, com sombreamento indicando períodos de baixa precipitação.

Interpretação: Períodos de Estiagem (barras cinzas): As barras cinzas marcam os períodos de seca.

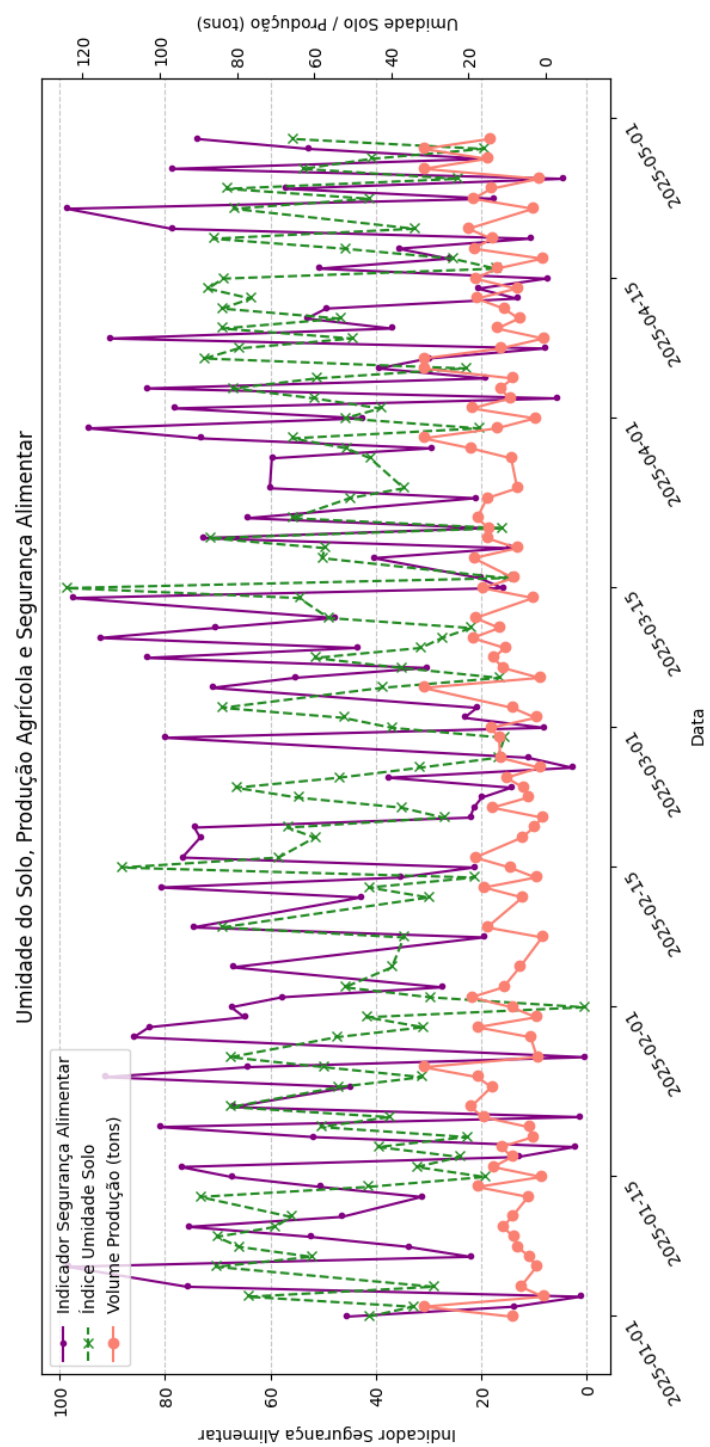


Figure 9: Gráfico de Séries Temporais: Índice de Umidade do Solo, Volume de Produção e Indicador de Segurança Alimentar

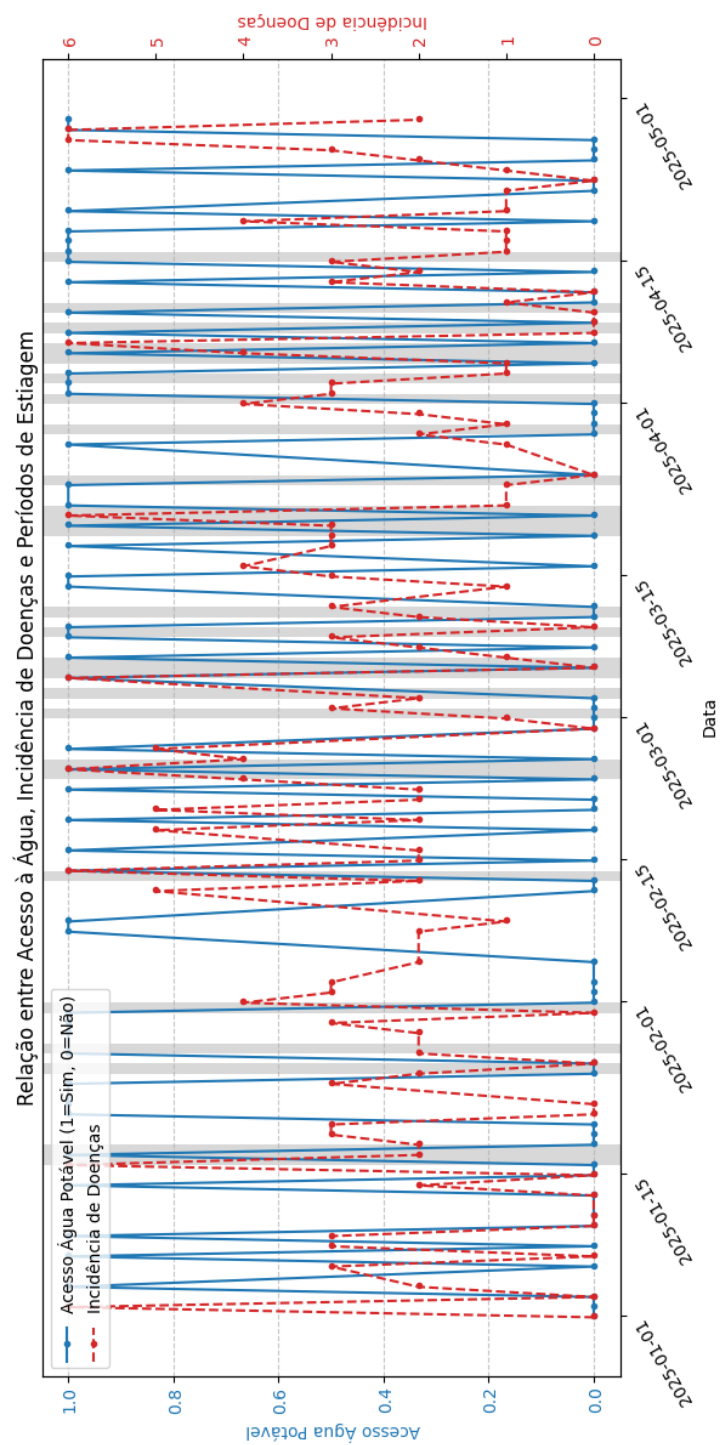


Figure 10: Gráfico de Séries Temporais: Acesso à Água Potável e Incidência de Doenças

Acesso à Água Potável (linha azul): O acesso à água potável (valor 1 = Sim, 0 = Não) é intermitente. O acesso é cortado (cai para 0) precisamente durante os períodos de estiagem marcados pelas barras cinzas.

Incidência de Doenças (linha vermelha tracejada): Assim como no Gráfico 2, a incidência de doenças tem picos agudos. Notavelmente, esses picos ocorrem exatamente durante ou logo após os períodos em que o acesso à água potável é interrompido.

Resposta: A análise sugere que existe uma relação direta entre a diminuição do acesso à água potável e o aumento da incidência de doenças, especialmente durante períodos de baixa precipitação. Durante os períodos de estiagem, o acesso à água potável diminui, o que está associado a um aumento na incidência de doenças.

5 Conclusão

A análise dos dados coletados e processados revelou insights significativos sobre os impactos das mudanças climáticas na agricultura local e na saúde da população. As respostas às perguntas centrais indicam que:

1. A região enfrenta uma forte estiagem que causa baixa umidade no solo, impactando diretamente e de forma negativa a produção agrícola.
2. As ondas de calor são um gatilho direto para o aumento de doenças na população, sugerindo a proliferação de doenças transmitidas por vetores (como dengue, que se beneficia do calor) ou problemas de saúde diretamente ligados à alta temperatura.
3. A crise hídrica e agrícola se traduz diretamente em uma crise de segurança alimentar, tornando a população vulnerável à escassez de alimentos.
4. A estiagem não só afeta a agricultura, mas também o abastecimento de água potável para a população. A falta de acesso à água segura está diretamente ligada a surtos de doenças, provavelmente de veiculação hídrica (cólera, diarreia, etc.) ou relacionadas à falta de higiene.

Essas descobertas ressaltam a necessidade urgente de intervenções para mitigar os efeitos das mudanças climáticas na região, incluindo estratégias de gestão de água, práticas agrícolas sustentáveis e programas de saúde pública para lidar com o aumento das doenças. A colaboração entre as comunidades locais, autoridades governamentais e organizações não governamentais será

crucial para enfrentar esses desafios e garantir a segurança hídrica e alimentar da população.

+ Aluno: Felipe Rafael Barbosa
+ Telefone: (91) 9 9244-0421